

Structural insights into mRNA recognition from a PIWI domain–siRNA guide complex

James S. Parker, S. Mark Roe & David Barford

Section of Structural Biology, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

RNA interference and related RNA silencing phenomena use short antisense guide RNA molecules to repress the expression of target genes^{1,2}. Argonaute proteins³, containing amino-terminal PAZ (for PIWI/Argonaute/Zwille) domains and carboxy-terminal PIWI domains, are core components of these mechanisms. Here we show the crystal structure of a Piwi protein from *Archaeoglobus fulgidus* (AfPiwi) in complex with a small interfering RNA (siRNA)-like duplex, which mimics the 5' end of a guide RNA strand bound to an overhanging target messenger

RNA. The structure contains a highly conserved metal-binding site that anchors the 5' nucleotide of the guide RNA. The first base pair of the duplex is unwound, separating the 5' nucleotide of the guide from the complementary nucleotide on the target strand, which exits with the 3' overhang through a short channel. The remaining base-paired nucleotides assume an A-form helix, accommodated within a channel in the PIWI domain, which can be extended to place the scissile phosphate of the target strand adjacent to the putative slicer catalytic site. This study provides insights into mechanisms of target mRNA recognition and cleavage by an Argonaute–siRNA guide complex.

RNA silencing is mediated by the effector complexes RISC (for RNA-induced silencing complex) and RITS (for RNA-induced initiation of transcriptional gene silencing), functioning in post-transcriptional and transcriptional silencing, respectively^{4–6}. Argonautes, comprising the Ago and Piwi subfamilies³, are the only proteins common to these complexes. Recent structural and biochemical studies implicate the PAZ domain in the recognition of an RNA 3' overhang^{7–10}. The PIWI domain is believed to incorporate the slicer catalytic site, which is responsible for the cleavage of target mRNA during RNA interference^{11–13}. However, these results

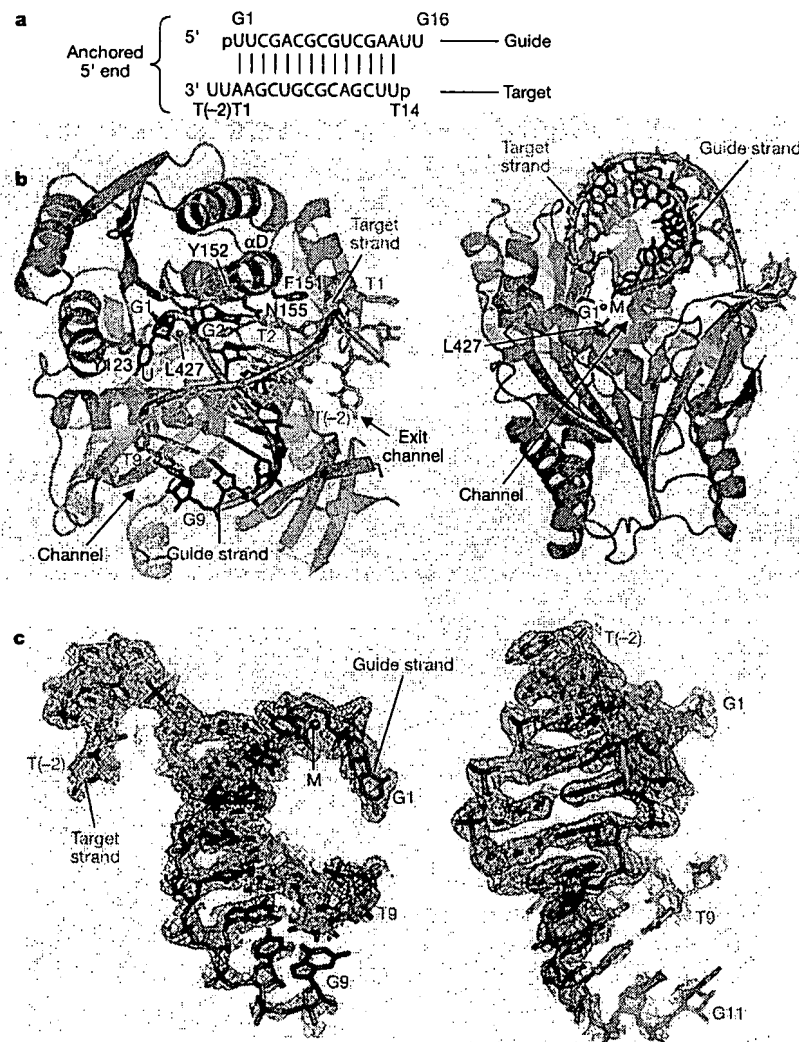


Figure 1 General features of AfPiwi–siRNA interactions and distortion of the 5' terminus of the RNA duplex. **a**, Diagram of the RNA duplex co-crystallized with AfPiwi, indicating the nomenclature for guide and target strands. **b**, Orthogonal views of the AfPiwi–RNA complex. The divalent metal in the G1-binding pocket is shown as a red sphere (M). **c**, Views of the RNA $2F_o - F_c$ electron density map contoured at 1σ . Left, view of the RNA as seen by the protein. Right, orthogonal view. G1 to G9 (G1 to G11 right) and T(–2) to T9 are placed in electron density. The figure was generated with PYMOL (<http://pymol.sourceforge.net>).

c, Views of the RNA $2F_o - F_c$ electron density map contoured at 1σ . Left, view of the RNA as seen by the protein. Right, orthogonal view. G1 to G9 (G1 to G11 right) and T(–2) to T9 are placed in electron density. The figure was generated with PYMOL (<http://pymol.sourceforge.net>).

are insufficient to explain the requirement for a 5' phosphate group on the first nucleotide of the guide strand^{13–16}, and the observation that the identity of the scissile phosphate in the target mRNA is measured relative to the 5' end of the guide strand, irrespective of modification or the length (within limits) of the 3' end^{15–18}.

AfPiwi, an isolated archaeal PIWI domain-containing protein, which serves as a model for understanding eukaryotic Argonaute PIWI domains, forms a single discrete complex with a 16-nucleotide siRNA-like duplex¹². With the aim of providing a molecular picture of PIWI domain–RNA interactions, we have crystallized and determined the structure of this complex at 2.2 Å resolution. We assign the two RNA strands as guide and target, denoting the guide nucleotides G1 to G16 (5' → 3') and complementary nucleotides T1 to T14 (Fig. 1a). The 5' end of the duplex refers to the 5' end of the guide.

AfPiwi binds to the 5' end of the siRNA-like duplex, engaging the RNA in an L-shaped channel (Fig. 1b). The RNA electron density is strongest at this end, allowing us to model nucleotides G1–G9 of the guide strand and T(–2)–T9 of the target (Fig. 1c). Nucleotides 2–9 on each strand base pair, adopting an A-form helix that extends along the channel (Fig. 1b and Supplementary Fig. 1). Strikingly, the

5' end of the RNA duplex is significantly distorted, resulting in the unwinding of G1 (uridine) and T1 (adenosine). G1, which carries the obligatory 5' phosphate of the guide, is displaced into a pocket surrounding the C terminus of the PIWI domain (G1-binding pocket). T1 passes through the short 'exit' portion of the channel, allowing the two 3' overhanging nucleotides, T(–1) and T(–2), to fold back along the protein surface. G1 and T1 are divided by the αD helix, which contributes two aromatic residues (Phe 151 and Tyr 152) capping the first 5' base pair of the RNA helix (Fig. 1b).

The G1-binding pocket is part of a highly conserved region in the PIWI domain, designated CRI (conserved region I)¹². A prominent structural feature is a divalent metal ion coordinated to the carboxylate group of the C-terminal residue, Leu 427 (Fig. 2a). This metal ion is crucial in the distortion and anchoring of the guide RNA, contacting the phosphates of G1 and G3. Without exception, all Argonaute sequences share three conserved hydrophobic residues at their extreme C terminus (Supplementary Fig. 2), indicating that an exposed carboxylate is an invariant structural feature.

Within the G1-binding pocket, the side chains of four invariant residues in Argonaute proteins (Tyr 123, Lys 127, Gln 137 and Lys 163; Supplementary Fig. 2) contact the 5' phosphate group of G1 (Fig. 2a). Tyr 123 further stabilizes the flipped conformation of G1 by stacking with the unpaired uracil base. In addition to interactions with these invariant side chains and the metal ion, the 5' phosphate contacts the backbone amide of Phe 138, fixed in position by the adjacent Gln 137. No other amino acids contact G1. Only five other residues are invariant in the PIWI domains of Argonaute proteins (Supplementary Fig. 2). Four of these seem to maintain the PIWI fold, and the fifth is a putative slicer catalytic site residue^{11,12}. The G1-binding pocket is therefore the most conserved structural feature of the PIWI domain. Consistent with our structural data is the observation that mutation of this pocket significantly decreased the affinity of AfPiwi for the siRNA-like duplex¹².

Interactions between AfPiwi and the guide are also mediated by the phosphate groups of G2–G5 (Fig. 2). A second key anchor site is the phosphate group of G3. In addition to its coordination by the metal, the phosphate interacts with Gln 159 and Arg 380, residues conserved in the Piwi and Ago subfamilies, respectively (Supplementary Fig. 2). The phosphate of G2 contacts Tyr 152, conserved as Tyr or Thr in the Ago subfamily, and the main chain amide of Arg 140. The phosphate of G4 also contacts Arg 380, and the phosphate of G5 interacts with Arg 385, through a water molecule. The only protein–base contact occurs between Asn 155, conserved in the Ago subfamily, and the 2' oxygen of the G2 uracil. Hydrogen bond acceptors occupy equivalent positions in cytosine and purines. Therefore, as expected, none of the interactions provide RNA sequence specificity. Interactions with G1–G5 account for all conserved residues of CRI (Fig. 3b), strongly indicating a conserved mode of RNA 5'-end binding within the Argonaute family. In contrast to the extensive interactions between the PIWI domain and the guide, few contacts are formed to the target strand, which is consistent with its requirement to dissociate after mRNA cleavage^{19,20}.

With G1 distorted into the conserved binding pocket, the corresponding base pair (G1–T1) is unwound (Fig. 1b). The finding that a single guide–target mismatch at this position can enhance slicer activity¹⁹ demonstrates that the AfPiwi–RNA complex corresponds to a cleavage-competent conformation. It also suggests an equilibrium between the preferred 'active' RNA duplex conformation that we observe and an 'inhibitory' conformation in which G1 pairs with T1. The mode of guide RNA binding to the PIWI domain—a highly ordered anchored 5' region, with bases exposed apart from G1—is consistent with the observation that for animal microRNAs, only nucleotides 2–8 of the microRNA guide (the 'seed') are critical for target recognition and pairing^{21–24}. Similarly, in RISC-mediated mRNA cleavage, bases at the 5' end of the siRNA guide dominate its affinity for target RNA¹⁹.

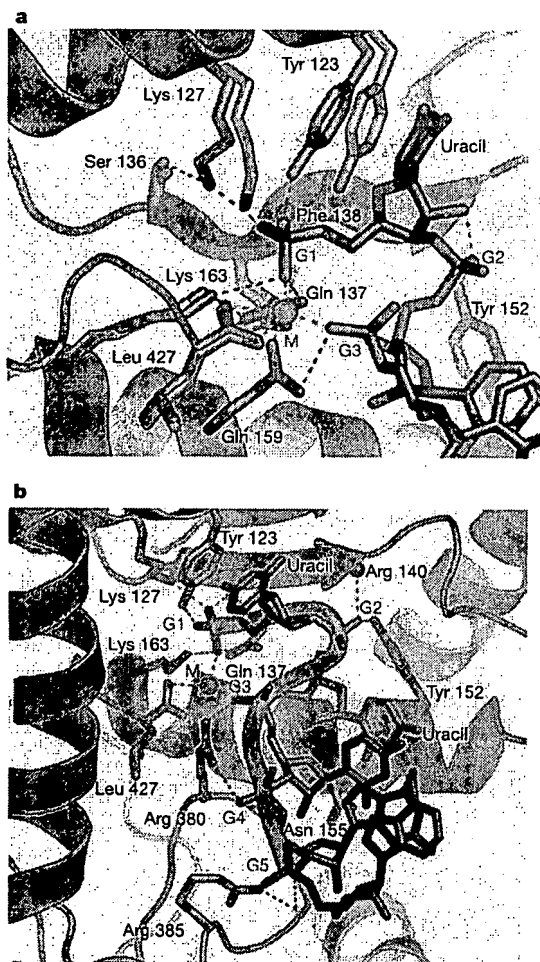


Figure 2 Interactions between AfPiwi and the guide RNA strand. **a**, Detailed view of the G1-binding pocket. Side chains coloured grey show conformations of selected residues in the absence of RNA¹². The concerted shift of Tyr 123 and Lys 127 suggests an induced-fit mechanism for phosphate binding. The divalent metal (M) is shown as a yellow sphere. **b**, View of interactions with the phosphate groups of G1 to G5. The target strand is omitted for clarity.

The G1-binding pocket and the putative slicer catalytic region^{11,12} lie at either end of the PIWI domain channel. Nucleotides G10–G16 of the guide and T10–T14 of the target are less clearly defined in our electron density maps. To visualize the trajectory of a full-length (19-nucleotide) siRNA guide bound to a mRNA target, we modelled the remaining nucleotides, assuming an A-form helix contiguous with the visible duplex (Fig. 3 and Supplementary Fig. 1). The weaker electron density for nucleotides G10 and G11 is compatible with this conformation (Fig. 1c). This model places the scissile phosphate of the target strand (linking T10 and T11)^{17,18} adjacent to the putative slicer catalytic region^{11,12} (Fig. 3 and Supplementary Fig. 3). The PIWI–RNA binding conformation is therefore consistent with our functional interpretation of the complex, being representative of a guide–target interaction, and provides insight into how the site-specificity of mRNA cleavage is

determined with reference to the 5' end of the guide^{17,18}.

Our structure might also reflect the initial phase of RISC association with a siRNA duplex; here the target strand would represent the passenger strand. The relative stabilities of the ends of an siRNA duplex determine the fate of the strands: the strand whose 5' end is located at the less stable end becomes the guide; the other (passenger) strand is discarded^{25,26}. Specific recognition of the siRNA-like 5' end, and partial unwinding of the duplex revealed from the Afpwi–RNA complex, indicates that binding to the PIWI domain might be the first step in siRNA duplex unwinding in RISC; PIWI could therefore participate in strand selection. Consistent with this notion is evidence supporting a role for Argonaute 2 in siRNA unwinding in *Drosophila*²⁷. R2D2 has also been implicated in guide strand selection, binding with a preference to the more stable end of an siRNA duplex²⁸.

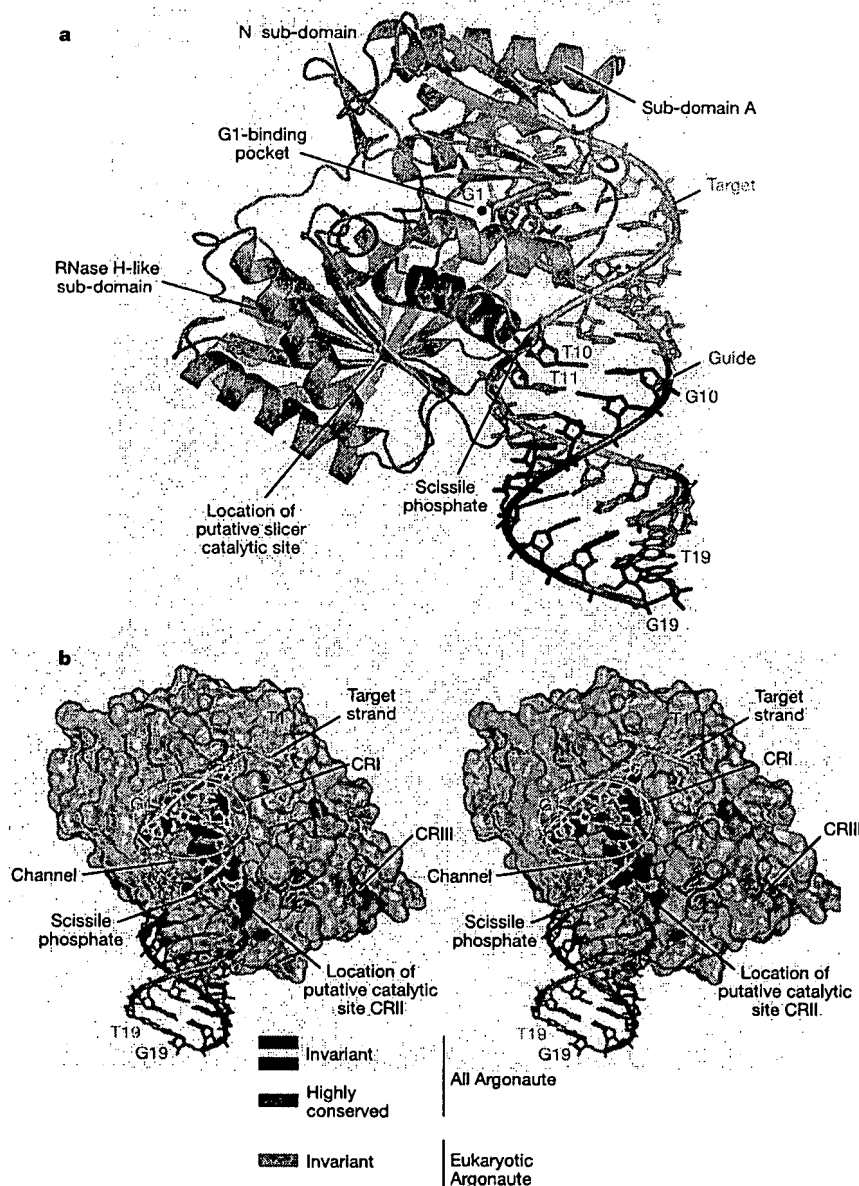


Figure 3 Proposed model for a 19-nucleotide guide–target duplex bound to Afpwi. **a**, View showing the positioning of the scissile phosphate of the target strand (between T10 and T11, coloured red) with respect to the location of the putative slicer catalytic site. The modelled RNA region is coloured darker for both strands (G10 to G19 and T10 to

T19). PIWI sub-domains are shown¹². Catalytic Asp residues of the DDE motif^{11,12} are absent in Afpwi. **b**, Stereo view with the molecular surface of Afpwi colour-coded according to structural conservation (Supplementary Fig. 2). Conserved regions CRI, CRII and CRIII are indicated¹².

A role for the PIWI domain in mediating interactions with the 5' end of the guide RNA complements the function of the PAZ domain in recognizing RNA 3' overhangs^{7–10}. In *Pyrococcus furiosus* Argonaute, the PAZ domain is positioned over the channel linking the G1-binding pocket and the putative slicer site¹¹. The PAZ domain may function as a mobile unit that locks down over the guide–target duplex, providing additional constraints for slicing¹¹. However, when bound to a target mRNA, it is unlikely that the recessed guide 3' end would interact with the PAZ domain¹⁰; instead, we envisage the duplex emerging into solvent or the immediate environment of RISC.

Our structure rationalizes many molecular and biochemical data for mechanisms of RNA silencing, which, together with the universal conservation of the G1-binding pocket, indicate that the AfiPiwi–RNA complex provides a representative picture of PIWI domain–RNA interactions. The view now emerges that Argonaute functions as a dynamic molecule mediating interactions with both the 5' and 3' ends of short guide RNA molecules, presenting them for recognition by cognate mRNA.

Note added in proof: In an accompanying paper³¹, Ma *et al.* use a different siRNA oligonucleotide to define the structure of an AfiPiwi–RNA complex. Their structure is similar to ours, although less of the RNA molecule is defined. Consistent with our model for 5'-end guide recognition in eukaryotic Argonaute proteins, mutation of the G1-binding pocket in hAgo2 ablates slicer activity. □

Methods

Complex formation and crystallization

AfiPiwi was purified as described¹², except that the protein was eluted at the final step in 5 mM HEPES pH 7.5, 800 mM NaCl, 2 mM MgCl₂, 1 mM MnCl₂, 8 mM dithiothreitol (DTT). The RNA oligonucleotide was purchased purified, deprotected and desalted from Dharmacon, and was resolubilized in 10 mM HEPES pH 7.5, 100 mM NaCl, 2 mM MgCl₂. Annealing was performed by incubation at 90 °C for 1 min followed by gradual cooling (1 °C per 5 min) from 65 °C to 4 °C. To form the complex, equal volumes of AfiPiwi (260 μM) and annealed RNA (310 μM) were mixed and incubated at 22 °C for 30 min. Crystallization was performed with the hanging drop method at 14 °C. Crystal clusters were obtained overnight by mixing 1 μl of complex with 1 μl of 200 mM KCl, 10 mM CaCl₂, 50 mM sodium cacodylate pH 5.5, 15% PEG 4000, 5 mM DTT. Single crystals were obtained by streak-seeding into a similar drop containing 12% PEG 4000 after incubation overnight. For cryoprotection, crystals were passed through stabilization solutions containing increasing amounts of 2-methyl-2,4-pentane diol (to 17%) and then flash-frozen at 100 K. Data were collected at ID14.EH1, ESRF.

Structure determination

The structure (with two AfiPiwi–RNA complexes per P1 unit cell) was solved by molecular replacement with PHASER²⁹ by using the native protein (PDB code 1W9H) as the search model¹². Some loops and all the RNA were rebuilt manually in COOT³⁰ and the complex was refined with REFMAC²⁹. A total of 508 water molecules were added in the final stages of refinement. The final *R* and *R*_{free} were 0.18 and 0.21, respectively (Supplementary Table 1). Terminal nucleotides of the target strands in the two RNA duplexes per unit cell adopt slightly different conformations (Supplementary Fig. 1). On the basis of crystallization conditions and refinement, we have assigned the metal ion within the G1-binding pocket as manganese.

Received 21 December 2004; accepted 17 February 2005; doi:10.1038/nature03462.

- Mello, C. C. & Conte, D. Jr. Revealing the world of RNA interference. *Nature* **431**, 338–342 (2004).
- Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
- Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
- Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
- Sontheimer, E. J. Assembly and function of RNA silencing complexes. *Nature Rev. Mol. Cell Biol.* **6**, 127–138 (2005).
- Yan, K. S. *et al.* Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468–474 (2003).
- Song, J. J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Struct. Biol.* **10**, 1026–1032 (2003).
- Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Struct. Biol.* **11**, 576–577 (2004).
- Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).

- Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
- Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727–4737 (2004).
- Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
- Nykanen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
- Chiu, Y. L. & Rana, T. M. RNAi in human cells: basic structural and functional features of small interfering RNA. *Mol. Cell* **10**, 549–561 (2002).
- Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
- Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877–6888 (2001).
- Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
- Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11**, 599–606 (2004).
- Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
- Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of *Drosophila* MicroRNA targets. *PLoS Biol.* **1**, 397–409 (2003).
- Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are MicroRNA targets. *Cell* **120**, 15–20 (2005).
- Khvorov, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
- Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
- Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
- Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
- Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Ma, J.-B. *et al.* Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* doi:10.1038/nature03514 (this issue).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to staff at ESRF for help with data collection. The work was funded by Cancer Research–UK, the Wellcome Trust and the Institute of Cancer Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B. (david.barford@icrac.uk). The coordinates and structure factors have been deposited with the Protein Data Bank with accession numbers 2bgg and r2bggsf, respectively.

Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein

Jin-Biao Ma¹, Yu-Ren Yuan¹, Gunter Meister², Yi Pei², Thomas Tuschl² & Dinshaw J. Patel¹

¹Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

²Laboratory of RNA Molecular Biology, The Rockefeller University, New York, New York 10021, USA

RNA interference (RNAi) is a conserved sequence-specific gene regulatory mechanism^{1–3} mediated by the RNA-induced silencing complex (RISC), which is composed of a single-stranded guide RNA and an Argonaute protein. The PIWI domain, a highly conserved motif within Argonaute, has been shown to adopt an RNase H fold^{4,5} critical for the endonuclease cleavage activity of